Deliverable 6.1

Project ID	654241
Project Title	A comprehensive and standardised e-infrastructure for analysing medical metabolic phenotype data
Project Acronym	PhenoMeNal
Start Date of the Project	1st September 2015
Duration of the Project	36 Months
Work Package Number	6
Work Package Title	PhenoMeNal Virtual Research Community Gateway
Deliverable Title	D6.1 User Experience document on VRC Design Guide
Delivery Date	M3
Work Package leader	EMBL-EBI
Contributing Partners	EMBL-EBI

Authors: Namrata Kale, Kenneth Haug and Pablo Moreno

Contents

1. Executive Summary 2
2. Project Objectives 2
3. Detailed report on the deliverable 3
a. Background 3
b. Workshop Overview 3
c. Introduction to the workshop 3
d. Workshop participants 4
e. Discussion and Outcomes 5
4. Delivery and Schedule 6
5. Background information 6
6. Annexes 7
Annex 1: Agenda 8
Annex 2: Workshop hand out 9
2.1 Introduction 9
2.2 Warm-up Activity 9
2.3 Activity A: Identify user groups 9
2.4 Activity B: Persona design for a key user group 10
2.5 Activity C: Experience map 11
2.6 Activity D: Identify and prioritise tools/services 12
2.7 Conclusion 13
Annex 3: Details of participants and work groups 14
Annex 4: Design document 16
4.1 Ranking of User Roles 16
4.2 Ranking of relevant Metabolomics-related Tools 17
4.3 VRC user personas 18
4.4 VRC opportunities 20
Annex 5: Photographs 23

1. Executive Summary

The PhenoMeNal Virtual Research Environment (VRE), aka. Virtual Research Community (VRC) will be the central access point for all users. This is the online portal enabling user interactions, virtual machine distribution, access to compute resources and information exchange. Users can search for existing public datasets and perform online analysis. The VRC will interact with the PhenoMeNal middleware to federate on-demand job requests between provisioned data, compute, and analysis resources. Expert help and training will be available through the VRC.

To ensure a user centric design and usability of the VRC, we arranged a User Experience (UX) workshop where we employed UX design principles like creating personas, experience mapping, storyboarding, card-sorting and poker prioritisation. We invited future VRC users from the consortia and potential users from the metabolomics community for a hands-on workshop to make sure the future VRC is designed to suite the users.

To ensure all project partners and collaborators adheres to the design standards, the UX VRC design guide will be distributed in the PhenoMeNal-h2020.eu website.

2. Project Objectives

No.	Objective	Yes	No
1	Arrange UX workshop	x	
2	D6.1 Document describing the user experience workshop on VRC design	x	
3	MS6.1 Release of VRC design guide (M3)	x	

3. Detailed report on the deliverable

a. Background

To ensure a user centric design, in other words a system that users would like to use, we applied well-established UX methods to ensure that the design is intuitive and users know how to interact with the VRC.

b. Workshop Overview

The workshop was conducted on 19th of November 2015 by EMBL-EBI at the James Watson Pavilion, Wellcome Trust Conference Centre, Hinxton, Cambridge. The detail of the agenda is provided in **Annex 1**.

The workshop included the basic elements of a User Experience Design: User design, interaction design, information architecture and other disciplines and involved participants from diverse backgrounds including bioinformaticians, clinicians, doctoral students, biostatisticians and software professionals. The User Experience Expert with the help of facilitators conducted a series of interactive activities involving active involvement of the participants.

c. Introduction to the workshop

The workshop was focussed on finding answers to some pressing questions addressed through this User Experience workshop (UX) and will eventually help setting up a user focused PhenoMeNal Virtual Research Community (VRC). The key questions identified were:

- Identify all users and the differences between them
- Identifying major users
- Gain an understanding of their current practices and workflow
- Identify the services they use
- Prioritise and group the services to integrate initially

Based on the needs to answer these questions, the workshop was split into four main activities:

- Activity A: Identify user groups.
- Activity B: Persona design for a key user group.
- Activity C: Experience map and user workflows.
- Activity D: Identify and prioritise tools/services.

A detail about the activities conducted during the workshop is attached as **workshop hand out in Annex 2.**

d. Workshop participants

A total of 23 participants including facilitators and an external UX expert attended the workshop. The facilitators included: Kenneth Haug, Pablo Moreno, Reza Salek and Namrata Kale from EMBL-EBI. Google Hangouts facilitated the online participation.

During the workshop, Christoph Steinbeck, co-ordinator of PhenoMeNal introduced the participants to the PhenoMeNal project and the aims of the e-infrastructure. The UX expert, Paula de Matos then introduced the goal of the workshop and the planned activities to achieve this goal. This was followed by the introduction of the participants.

The participants were divided into 4 groups and were assigned to a facilitator. The group division was made according to the technical abilities and qualifications ensuring fair representation of users from different backgrounds. A detail of the participants with the facilitator is included in **Annex 3**.

e. Discussion and Outcomes

The workshop produced three essential outputs:

- Ranking of relevant user roles (aka. personas) towards which the VRC should be initially designed (user roles). The most common users identified were:
    - Bioinformaticians
    - Clinicians in research/clinical data analyst
    - Principal Investigators (PIs) / scientists not actually working in wet labs
    - Scientist in the bench

  Further details regarding how this ranking was achieved can be found in **Annex 4, part 4.1.**
- An experience map and practical user workflows for some of these relevant user roles.
- A prioritised list of relevant existing software tools we should containerise and make available through the VRC. The most relevant Metabolomics-related Tools that were identified to be integrated are:
    - XCMS (https://bioconductor.org/packages/release/bioc/html/xcms.html)
    - MetaboLights (http://www.ebi.ac.uk/metabolights/)
    - OpenMS (http://open-ms.sourceforge.net)
    - ROPLS (https://www.bioconductor.org/packages/3.3/bioc/html/ropls.html)
    - msConvert (http://proteowizard.sourceforge.net/tools.shtml)
    - Galaxy (https://galaxyproject.org)
    - R (https://www.r-project.org)
    - CAMERA (http://msbi.ipb-halle.de/msbi/CAMERA)

  Further details regarding how this ranking was achieved can be found on **Annex 4, part 4.2**. It is worth mentioning that we expect to find more tools to be relevant by further conversations with Metabolomics analysts in different groups. Compared to the roles presented before, the tools ranking is less relevant as many more tools (than roles) will be considered for the system.

In line with the various discussions following the relevant activities conducted during the workshop, it was concluded that all software, virtual images and other containerised components should be made available for any anonymous users/occasional-browsing individuals to download and use locally. Registered users should in addition be able to register and upload new and changed components.
It was also suggested that all software incorporated or developed during the PhenoMeNal project has to be 100% open source and be free to use for any purpose. Existing software that is included as-is, will be governed by the respective license agreement already in place. All consortium members will continue to apply these techniques in the future development of the VRC and any new modules deployed within. In practical terms this means that we will continue to build upon the

personas and prioritised lists we developed in this workshop and, when required, interview prospective user groups to gain further domain knowledge.

4. Delivery and Schedule

The delivery is delayed: No

5. Background information

The PhenoMeNal Virtual Research Community (VRC) will be the central access point for all users. This is the online portal enabling user interactions, virtual machine distribution, access to compute resources and information exchange. Users can search for existing public datasets and perform online analysis. The VRC will interact with the PhenoMeNal middleware developed in WP5 to federate on-demand job requests between provisioned data, compute, and analysis resources. Expert help and training will be available through the VRC.

Work package number	WP6	**Start date or starting event:**	M1
Work package title	PhenoMeNal Virtual Research Community Gateway		
Participants	EMBL-EBI, ICL, INRA		

Objectives

1. **6.1** Establishment of the PhenoMeNal Virtual Research Community (VRC) portal

Task 6.1: User Experience workshop on VRC design (**EMBL-EBI,** ICL, INRA)
To ensure a user centric design and usability of the VRC, we will have a workshop where we employ usability design principles like storyboarding, card-sorting and low-fidelity mockups (website drafts). We will invite potential VRC users from the community for a hands-on workshop to make sure the VRC is designed to suite the users.

6. Annexes

1. Agenda
2. UX meeting hand out and plan of the day
3. Details of participants and work groups
4. Design document
5. Photographs

Annex 1: Agenda

9:30 - 9:50	**Welcome, Introduction and Warm-up activity.**
9:50 - 10:30	Activity A: Identifying user groups of VRC.
10:30 - 11:00	**Coffee break**
11:00 - 11:15	Activity A: Identifying user groups of VRC cont.
11:15 - 11:45	Activity B: Persona design for key user groups.
11:45 - 12:00	Activity B: Persona design report back.
12:00 - 12:30	Activity C: Experience mapping: Understanding user workflows and tasks for these user groups.
12:30 - 13:30	**Lunch**
13:30 - 14:00	Activity C: Experience mapping contd.
14:15 - 14:30	Activity C: Experience map group report back.
14:30 - 15:00	Activity D: Identify services and tools to prioritise.
15:00 - 15:30	**Coffee break**
15:30 - 16:00	Activity D: Planning poker to prioritise the categories.
16:00 - 16:15	Activity D: Tools and services report back.
16:15 - 16:30	Closing remarks

Annex 2: Workshop hand out

2.1 Introduction

Welcome to the UX Workshop where we will answer and identify key questions about our users. Today we will do four activities over the course of the day.

- Activity A: Identify user groups.
- Activity B: Persona design for a key user group.
- Activity C: Experience map and user workflows.
- Activity D: Identify and prioritise tools/services.

2.2 Warm-up Activity

Lets go round the room and make some introductions. Please answer these 3 questions when it's your turn.

1. What's your name?
2. Where do you work?
3. One thing about you that nobody in this room knows?

2.3 Activity A: Identify user groups

In this session we are aiming to identify all the user groups who could use this VRC portal and try to understand their needs.

1. On post-it notes write down all user groups you can think of. Do this individually first – don't discuss with your group at this point. Your group should all use the same colour post-it notes for this part. (5 minutes)
2. Once you have written all the user groups you can think of, place all the post-it notes on the foam board. As a team then try to organise the user groups into related groups. For example, post-it's, which are identical, should be grouped together. Try to label the categories for each user group. Use a different colour post-it note for categories. (~30 minutes)
3. Now that you have got all the user groups, write them down in a list. We now going to play a game called planning poker. Now give each individually a planning poker set of cards labelled 1 to 5. Read out the name of the user group and let everyone assign a number of importance of that user group i.e. 5 being most important and 1 being least important. Everyone should place their card face down and reveal them all at the same time. If the numbers are vastly different it's worth having a discussion about it. Add up the card numbers and assign the total to the user group. See the table below as an example.

User group	Total
(e.g. plumber)	10 (5+3+2)
(e.g. dishwasher)	7 (2+3+4)

At the end of this activity you should have an ordered list of your top priority user groups. You should also have a wide variety of user groups that you perhaps had not thought of.

2.4 Activity B: Persona design for a key user group

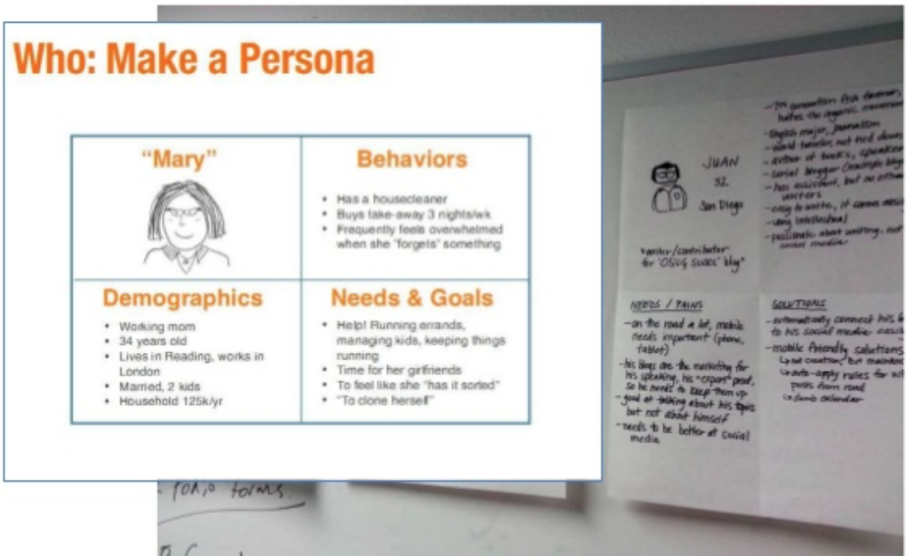
Of your top 3 or 4 user groups prioritized in the previous activity, pick **one** to develop further. In this activity we will design an archetype user group based on facts we know about them. The goal is to create somebody we can empathise with.

As a group use the A3 flipchart to write down your persona. Divide the chart into six sections and label them as below. For each section discuss and fill in the details. If you wish you can also do this online on Google docs for online participants.

As a group use the A3 flipchart to write down your persona. Divide the chart into six sections and label them as below. For each section discuss and fill in the details. If you wish you can also do this online on Google docs for online participants.

Name Picture "*One line quote that describes this user*"	Behaviours
Demographics	Needs & Goals
Pain Points *"Issues they have in their day to day work which the VRC can potentially help with."*	Scenario of use *"Describe a scenario from the perspective of this user of how they would use the VRC."*

Here is an example proto-persona with just 4 sections.

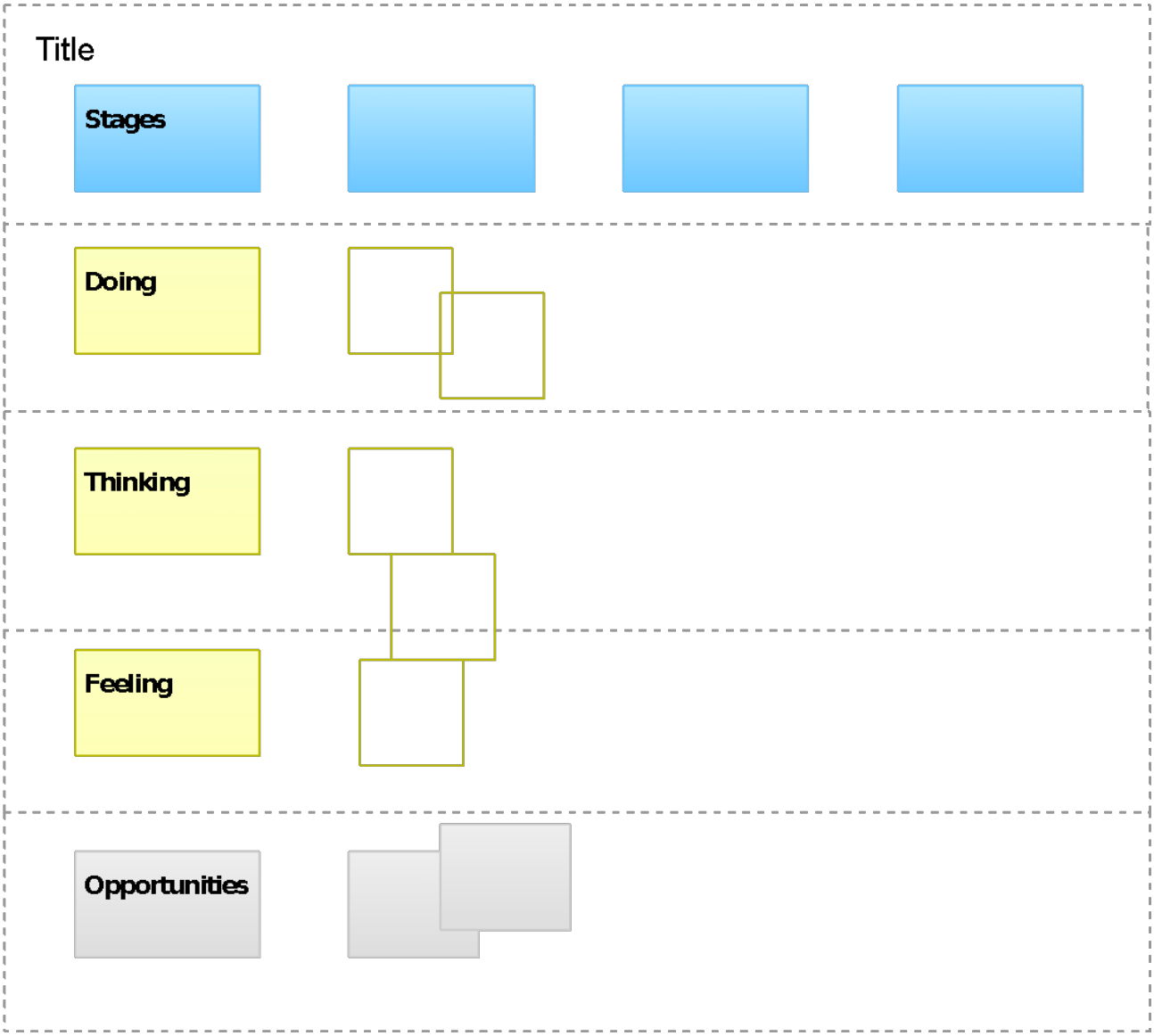


2.5 Activity C: Experience map

In this exercise we will go into more detail of how the persona will use the VRC. Use the scenario you described in the previous section as the starting point for this activity.

You can see an example template (next page) that you will be using for this activity. You can use a foam board for this activity and stick on the post-it notes to label each section. Remember to give your Experience Map a title. Below follows a description of each section:

- Stages – describe the distinct stages of the scenario. It is highly likely that this could non-linear but you don't have to go to immense detail but give the stage a title.
- Doing – here you want to describe exactly what is happening at this stage. It could be shown as a flowchart.
- Thinking – what is our persona thinking about at this stage.
- Feeling – what are they feeling e.g. frustrated, satisfied.
- Opportunities – here you should describe the opportunities the VRC has which can impact



2.6 Activity D: Identify and prioritise tools/services

The aim of this activity is to understand all the tools and services that the community is using. Then we want to prioritise which ones should be integrated first.

This is a similar activity to Activity A but instead we are focussing on tools/services.

1. On post-it notes write down all user groups you can think of. Do this individually first – don't discuss at this point with your group. Your group should all use the same colour post-it notes for this part. (5 minutes)
2. Once you have written all the tools/services you can think of, place all the post-it notes on the foam board. As a team then try to organise the post-its into related groups. For example, post-it's, which are identical, should be

grouped together. Try to label the categories for each group. Use a different colour post-it note for categories.

3. Now that you have got all the groups, write them down on an online table. Now give each individually a planning poker set of cards labelled 1 to 5. Read out the name of the user group and let everyone assign a number of importance of that user group i.e. 5 being most important and 1 being least important. Everyone should place their card face down and reveal them all at the same time. If the numbers are vastly different it's worth having a discussion about it. Add up the card numbers and assign the total to the user group. See the table below as an example.

User group	**Total**
(e.g. plumber)	10 (5+3+2)
(e.g. dishwasher)	7 (2+3+4)

2.7 Conclusion

We have now completed all the activities and it should be much clearer who our users are, what they are trying to do and what services should be prioritised.

Annex 3: Details of participants and work groups

External UX Expert: Paula de Matos

Group: 1. **Facilitator**: *Ken Haug (EMBL-EBI)*

S.no	Name	Affiliation	Expertise	Participation
1	Chris Steinbeck	EMBL-EBI	Experimentalist/ Programmer	In person
2	Kim Kultima	UU	Experimentalist	In person
3	Dario Vianello	EBI	Cloud bioinformatics application architect	In person
4	Michael van Vliet	University of Leiden	Scientific programmer	Online

Group: 2. **Facilitator**: *Pablo Moreno (EMBL-EBI)*

S.no	Name	Affiliation	Expertise	Participation
1	Xuefei Li	MRC/EBI	Experimentalist	In person
2	Payam Emami	UU	PhD Bioinformatics	In person
3	Gianni Dalla Torre	EBI	Programmer	In person
4	Tim Ebbels	ICL	Experimentalist	Online

Group: 3. **Facilitator**: *Reza Salek (EMBL-EBI)*

S.no	Name	Affiliation	Expertise	Participation
1	Rachel Spicer	EMBL-EBI	PhD student Bioinformatics	In person
2	Daniel Schober	IPB	Ontologist	In person
3	Jose Dianes	EMBL-EBI	Software Engineer	In person
4	Antonio Rosato	CIRMMP	Experimentalist	In person
5.	Vitaly Selivanov	University of Barcelona	Bioinformatician	Online

Group: 4. **Facilitator**: *Namrata Kale (EMBL-EBI)*

S.no	Name	Affiliation	Expertise	Participation
1	Etienne Thevenot	CEA	Biostatistician	In person
2	Jose Ramon	EMBL-EBI	Bioinformatician	In person
3	Ibrahim Karaman	ICL	Chemometrician	In person
4	Venkata Chandrasekhar	EMBL-EBI	Software Engineer	In person
5	Pedro de Atauri	University of Barcelona	Bioinformatician	Online

Annex 4: Design document

4.1 Ranking of User Roles

The following table provides the collated results for Roles identified, and ranks them according to combined criteria of all the User Experience groups. A User Experience group is a set of the participants as shown in Annex 3 of this document. Roles are ordered descending in order of importance.

User Role	Frequency Ranking	Avg. Rank Ranking	Combined Ranking
Bioinformatician	1	2	3
Clinician in research/clinical data analyst	3	1	4
Scientist not in the bench, planning	3	3	6
Scientist in the bench	1	5	6
Clinician in clinic and not in research	3	6	9
Chemometrician	10	4	14
Teacher/Trainer	7	7	14
IT Department/IT Infrastructure manager	3	13	16
Integrative omics	10	7	17
Non-clinical application domains	10	7	17
IT Applications/Software developer	7	10	17
Funding Agent-Administrator	7	12	19
Audit	10	11	21

Frequency Ranking: Considers how many of the UX groups identified this Role. The more groups that identify a particular Role, the higher its ranking (1 is the highest ranking).
Avg. Rank Ranking: Within each UX group, each Role was ranked (when identified). This is simply the average out of all those rankings.
Combined Ranking: Takes into account, with the same weight, both previous ranking, to hopefully highlight Roles that were identified many times with high rankings amongst each UX group.

4.2 Ranking of relevant Metabolomics-related Tools

This table provides the collated results for Tools identified, and ranks them in descending order according to combined criteria of all the User Experience groups. A User Experience group is a set of the participants as shown in Annex 3 of this document.

Tool	Frequency Ranking	Avg. Rank Ranking	Combined Ranking
XCMS	1	7	4.5
MetaboLights	1	9	5.5
OpenMS	4	4	6
ROPLS	7	1	7.5
msConvert	7	1	7.5
Galaxy	7	1	7.5
R	7	4	9
CAMERA	7	4	9
Other Reference Spectra DB	3	13	9.5
Statistics	7	8	11
MetFrag	4	14	11
rNMR	7	10	12
BiNChE	7	10	12
MetExplore	7	12	13
ISA-Tools	4	18	13
MetFusion	7	14	14
ipython	7	14	14
IPO	7	14	14
MetaboAnalyst	7	18	16
BATMAN	7	18	16
Proteowizard	7	18	16
Pathway Visualisation	7	22	18
R-Normalization	7	22	18
R-plotting packages	7	24	19
pyOpenMS	7	25	19.5
MZML	7	25	19.5
NMRML	7	27	20.5
Workflow4Metabolomics	7	28	21

Frequency Ranking: Considers how many of the UX groups identified this Tool. The more groups that identify a particular Tool, the higher its ranking (1 is the highest ranking).
Avg. Rank Ranking: Within each UX group, each Tool was ranked (when identified). This is simply the average out of all those rankings.
Combined Ranking: Takes into account, with higher weight for Avg. Rank Ranking, both previous ranking, to hopefully highlight Tools that were identified many times with high rankings amongst each UX group.

4.3 VRC user personas

Each group mapped out all possible future VRC users the members could think of. As a team these users were organised into related groups. To focus the discussions and identifying common workflows and future opportunities, each group focused on one "persona". This persona describes a fictitious person, in a real life job, that is a potential future user of the VRC.

Group	Persona name and short description
1	***Kristina***, Clinical chemist, female, 32 years old **Needs & Goals**: Minimise manual labour, Automation, Quality control, High throughput / Volume **Scenario of use**: ● Automated/pipelined methods ● 1000 urine samples per week ○ 3 mins per sample for each data file ○ Data files has to go through automated QC ○ Targeted ○ All against internal samples ○ Results, minimise manual checking
2	***Bea***, Bioinformatician, female, mid 30's, BSc **Needs & Goals:** Needs data, fancy workspace, collaborator who produces data, flexible work time, hardware and connectivity, publish results, find public data, find software, reproduce research, funding for next years **Scenario of use:** ● Has MS data from own group that needs to be analysed (data collection, initial quality control (QC), open format conversion, pre-processing, analysis, publish).

	• University network is slow and has little storage space, yet increasing amounts of data. • Has knowledge of tools to use, but not necessarily the time to spend on installing them all “just for trying them”. Would love to try them without this burden. • On University hardware, most of the analysis takes many hours to run, and hence failures make the process extremely slow. • Once her pipelines are set, she would love to control/deploy/monitor them programmatically (REST API?), possibly remotely. • Pipelines need to be run for many different data sets. • Needs to be able to observe intermediate outputs for monitoring and QC purposes. • She is good at the scientific computing side of things, possibly not so much at the software engineering level. • Concern about portability of her pipelines. • Wants to avoid repetitive tasks.
3	***Alex***, Professor of metabolomics, female, 40 years old **Needs & Goals**: Standardisation for reproducibility, transferable, student engagement, Open Source **VRC Scenario of use**: • Training for end-users ○ e.g. medics, biologist, bioinformaticians • Beginners and advanced
4	***Anne***, Chemometrician, female, 30 years old **Needs & Goals**: Implement own tools, reproducibility, likes benchmarking, not computer scientist so need simplification, needs data and documentation, reproduce data **VRC Scenario of use:** • Cross-datasets & networks • User search interface • Quality control • Normalisation • Feature selection • Classification • Experimental design

4.4 VRC opportunities

Each group in the workshop decided on a common daily workflow for their respective personas, e.g. a set of tasks this person would perform in a regular bases, and also how the respective persona relates to these tasks. Based on this input, we identified the main opportunities in which the VRC can impact in given each scenario.

User	Stage ● Opportunities
Kristina, the Clinical Chemist	Run order creation ● Robotisation (automatisation) Metabolite quantification ● Open source non-vendor specific quantification software. QC (using vendor software) ● QC on the fly. ● Parse patient journal details to target only the requested compounds. QC (visual inspection) ● Statistics. ● Machine learning. Clinical Decision Support ● Teaching Medical Doctors.
Bea, the Bioinformatician	Collecting Data ● Data flow could be done over the network. Initial QC ● Auto-check run order. ● Checksum checks. ● Reach collaborators for fixing things through the VRC. ● Experimental design file mapping. ● UIDs for files and samples. Conversion to Open Format ● This is mostly local to data producer; vendor libraries required. ● Reduce transfer burdens. ● Be able to estimate times. Pre-processing ● Support pre-processing pipeline.

	• Improve communication with collaborator. • Automate parameter estimation and automatisation. • Share parameters among similar technologies/settings. • Data visualisation for spectra. • Comparison of multiple pipelines for pre-processing. Analysis • Pipelines for all parts. • Provide plenty of tools. • Visualization of outputs. • Feed results directly to collaborators. • Faster processes on cloud. Publish • Publication ready figures. • Integrated data reduction. • Reproducible research.
Alex, the Prof. of Metabolomics	Set-up course • Funding. • Support. • Dockers available • Teaching. Setting up program • Help with IT from PhenoMeNal Training sessions • Improving tools, material, processes. Feedback • Feedback on funding.
Anne, the Chemometrician	Data Import • Well prepared data, detailed checking • Audit reports QC (Bad samples) • New metrics, visualization. • Audit reports. Pre-process • Combine and customize latest techniques. • Default parameters, pre packaged workflows, audit reports. QC Outliers • Reporting which samples are removed and why.

	• Audit report. Statistics • Robust, complimentary and up to date methods. • Audit report. Visualization and Reporting • Direct import to MetaboLights, report generation, audit and export back.

Annex 5: Photographs

Photo 1: Workshop Participants assemble in groups for a brainstorming session on defining user groups with the UX expert Paula

Photo 2. Facilitators from each group summarise the outcome of the brainstorming session on user groups. Other participants are following remotely via google hangout